Journal of Experimental Psychology:
Learning, Memory, and Cognition
2012, Vol. 38, No. 4, 1019–1029

0278-7393/12/$12.00 DOI: 10.1037/a0026976

A Processing Approach to the Working Memory/Long-Term Memory Distinction: Evidence From the Levels-of-Processing Span Task

Nathan S. Rose and Fergus I. M. Craik
Rotman Research Institute of Baycrest, Toronto, Ontario, Canada

Recent theories suggest that performance on working memory (WM) tasks involves retrieval from long-term memory (LTM). To examine whether WM and LTM tests have common principles, Craik and Tulving's (1975) levels-of-processing paradigm, which is known to affect LTM, was administered as a WM task: Participants made uppercase, rhyme, or category-membership judgments about words, and immediate recall of the words was required after every 3 or 8 processing judgments. In Experiment 1, immediate recall did not demonstrate a levels-of-processing effect, but a subsequent LTM test (delayed recognition) of the same words did show a benefit of deeper processing. Experiment 2 showed that surprise immediate recall of 8-item lists did demonstrate a levels-of-processing effect, however. A processing account of the conditions in which levels-of-processing effects are and are not found in WM tasks was advanced, suggesting that the extent to which levels-of-processing effects are similar between WM and LTM tests largely depends on the amount of disruption to active maintenance processes.

Keywords: short-term memory, working memory, long-term memory, secondary memory, depth of processing, levels of processing

The idea that short-term memory (STM) and long-term memory (LTM) represent distinct memory systems has a long history (see Jonides et al., 2008, for an excellent review). Memory over the short term and the long term has been thought to differ in many ways in terms of capacity, the underlying neural substrates, and the types of processes that support performance. Following the ideas proposed by Baddeley and Hitch (1974), studies in the general area of short-term retention have been increasingly framed in terms of working memory (WM), typically involving tasks in which cognitive operations are performed on small amounts of information held briefly in mind. Although there is general agreement that short-term retention tasks rely on STM or WM and longer term retention tasks rely on LTM, there is less agreement on the cognitive architecture of STM, WM, and LTM and on the relations between these arguably distinct types of memory. The situation is further complicated by the possibility that short-term retention tasks may draw on information held in both short-term and long-term memory. For example, Waugh and Norman (1965) proposed that the recency effect in verbal free-recall reflects retrieval from both *primary memory* (PM) and *secondary memory* (SM). Many current researchers (e.g., Cowan, 1999; Oberauer, 2002; Unsworth & Engle, 2007) also believe that short lists of items (e.g., four words) can be maintained in PM (or the *focus of attention*; Cowan, 1999), whereas recalling longer lists and items from complex WM tasks additionally involves retrieving items that have been displaced from PM, and so these items must be retrieved from LTM (or SM).

The purpose of the present article is to elucidate the conditions under which performance on a WM task reflects retrieval from both PM and SM and thereby contribute to the debate about the nature and architecture of the WM "system." The terminology in this area is admittedly confusing, and exacerbated by the possibility that performance on STM tasks and WM tasks may reflect retrieval from more than one system. As far as possible we use the terms PM and SM to refer to theoretical constructs, and the terms STM, WM, and LTM to describe different types of tasks.

On the Distinction Between WM and LTM

Several lines of evidence support the idea that the characteristics of memory over the short and long term are substantially different. One concerns the difference in the amount of information that can be maintained. PM is limited in capacity in that only a finite number of items can be maintained in conscious awareness at any given time; in contrast, the capacity of SM is assumed to be virtually limitless. Another source of evidence for the existence of two distinct systems is provided by cases of amnesia following brain damage. Damage to the hippocampus produces an inability to form or retrieve new long-term memories, but the ability to maintain and reproduce a small subset of information over the short term is typically preserved (Baddeley & Warrington, 1970).

Nathan S. Rose and Fergus I. M. Craik, Rotman Research Institute of Baycrest, Toronto, Ontario, Canada.

This research was conducted as part of a dissertation submitted by Nathan S. Rose to the Washington University Graduate School of Arts and Sciences.

We thank Joel Myerson, Roddy Roediger, Sandy Hale, and Mitch Sommers for their guidance and support and for their comments on an earlier version of this article. We also thank Denise Rose and Matt Robbins for their support and assistance with data collection and scoring. Finally, we owe a debt of thanks to David P. McCabe, who contributed to an earlier version of this article before his death in 2011.

Correspondence concerning this article should be addressed to Nathan S. Rose, The Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario, M6A 2E1, Canada. E-mail: nrose@rotman-baycrest.on.ca

In contrast, patients with damage to perisylvian cortex, such as patient KF, show the reverse pattern of impairment: preserved performance on LTM tasks, but impaired performance on STM tasks (Shallice & Warrington, 1970). However, more recent studies have raised questions concerning this double dissociation (Jonides et al., 2008; Ranganath & Blumenfeld, 2005; Rose, Olsen, Craik, & Rosenbaum, 2011).

A further point of distinction between PM and SM concerns differences in the type of encoding, maintenance, and retrieval processes involved in performance on tasks thought to tap the two systems. For example, on STM or WM tasks that require remembering a series of words, people tend to rehearse the words, and their performance is better when they can do so without distraction. In contrast, on LTM tests, it is usually not possible to rehearse a long list after only a single presentation or to continuously rehearse even a short list over a long delay. Instead, people perform better on explicit LTM tests when deeper (semantic) cues are encoded at the time of initial learning than when shallower (perceptual) cues are encoded—the so-called levels-of-processing effect (LOP; Craik & Tulving, 1975). Evidence that verbal STM tasks rely heavily on phonological or articulatory codes, whereas LTM tasks typically involve semantic codes, was presented by Baddeley (1966a, 1966b) and was developed further in the multicomponent model of WM (Baddeley, 1986; Baddeley & Hitch, 1974). Given this difference in representational codes that usually underlie performance in short-term and long-term tasks, it may be preferable to talk directly about these codes and their characteristics, as opposed to differences in putative systems. We develop this line of argument later in the article.

Historically, the concept of WM may be thought of as evolving out of the concept of STM or PM. Whereas STM was assumed to be devoted solely to the temporary storage of information, the concept of WM was developed to capture more dynamic processes to explain performance on tasks requiring the simultaneous engagement of processing activities in addition to temporary storage. For example, something more than just temporary storage of information is clearly needed to perform complex cognitive activities such as language comprehension, mathematics, and reasoning. As the concept of WM developed, however, there was a theoretical shift in the way researchers conceptualized the distinction between WM and LTM. Whereas the distinction between STM and LTM was once quite clear, the division between WM and LTM is considerably less well specified.

The original multicomponent model of WM (Baddeley, 1986; Baddeley & Hitch, 1974) included a central executive component responsible for controlling attention during the performance of a task, in addition to a set of temporary storage buffers dedicated to the maintenance of particular types of information, but this original model did not address a number of empirical and theoretical issues concerning the ways in which WM and LTM are related—issues that Baddeley (2007) recently described as "skeletons in the working memory cupboard" (p. 141). For example, Baddeley's (1986) model retained a distinct separation between WM and LTM, but this separation fails to account for why maintaining familiar information (e.g., words or symbols that already have representations in LTM) is easier than maintaining novel information (e.g., nonwords or novel shapes). As a second example, span level for a series of unrelated words is typically around five to six words, but if the words form a coherent sentence, span rises to 15–20 words. Such examples unequivocally demonstrate a contribution of long-term (semantic) memory to WM performance (cf. Thorn & Page, 2008). To account for such effects, Baddeley (2000) added a further component called the episodic buffer, another store to which a variety of multicode, task-relevant information is brought from LTM. Baddeley (2000) now argues that an important function of this new component is to act as an interface between the other buffers of his multicomponent model and the LTM system.

Other models of WM posit that information "in WM" is embedded within LTM but that this information is highly activated and easily accessible. For example, Cowan's (1999, 2005) embedded process model proposed that items in long-term (secondary) memory may exist in varying states of accessibility based on their level of activation. Items that have been recently perceived or processed, or that are currently being attended to, are activated representations. According to Cowan's model, only a small number (i.e., four) of such activated representations or "chunks" can be held within a component termed the "focus of attention" (a concept that is essentially synonymous with PM; N. Cowan, personal communication, June 13, 2008). Oberauer's (2002) model of WM is similar to Cowan's in that memory items may exist in varying states of accessibility. Recently processed items have the most activated representations and are immediately accessible. However, according to Oberauer, only one item or chunk—not four—may be in the focus of attention at any given time. Nevertheless, Oberauer has acknowledged that recently activated items or chunks (items in a "region of direct access") may have privileged accessibility relative to other items in LTM. McElree (2006) also viewed focal attention as having a maximum capacity of one item and recall of other items involving retrieval from LTM. Most importantly for current purposes, all these researchers agree that items maintained in WM represent an activated subset of long-term (secondary) memory.

Unsworth and Engle's (2007) recent dual-component model of WM also suggests a large amount of overlap between WM and LTM tasks. Notably, Unsworth and Engle have reintroduced the concepts of PM and SM to the debate by proposing that WM relies on both components. That is, a small number of items (e.g., one to four) may be simultaneously maintained within PM, but when PM capacity has been exceeded, retrieval from SM is required, even though the time between encoding and retrieval is much shorter than in traditional LTM tasks. Unsworth and Engle have proposed that different types of immediate recall tasks engage PM and SM to varying degrees. For example, simple span tasks (sometimes called STM tasks), such as digit span, capture the ability to maintain a list of items and report them directly from PM. In contrast, complex span tasks (sometimes called WM span tasks), such as reading span or operation span, require participants to perform a secondary processing task (e.g., reading sentences, solving math problems, etc.) interleaved between presentation of to-be-remembered items. According to Unsworth and Engle's dual-component model, such secondary tasks require that participants temporarily switch attention away from maintaining the to-be-remembered items in PM. Therefore, although a few items may be reported from PM, at least some of the items must be retrieved from SM. Thus, according to Unsworth and Engle, although both PM and SM are involved in performing both simple and complex span tasks, the simple span tasks rely much more on PM, while complex span tasks rely for the most part on SM. This hypothesis is consistent with findings of enhanced long-term retention for items initially recalled from complex as opposed to simple span tasks (McCabe, 2008; Rose, 2011) and longer versus shorter lists (Rose, Myerson, Roediger, & Hale, 2010).

Other researchers suggest that not all secondary tasks displace information from PM to the same degree. For example, building on Johnson's (1992) description of refreshing ("thinking briefly of a just-activated representation"; Raye, Johnson, Mitchell, Greene, & Johnson, 2007, p. 135), the time-based resource-sharing model of Barrouillet, Bernardin, and Camos (2004) assumed that participants attempt to refresh to-be-remembered items between the processing and encoding phases of complex span tasks, and this helps to maintain the items so that they are accessible at the time of recall. Compelling evidence has been presented to suggest that refreshing is distinct from rehearsal (Barrouillet et al., 2011; Raye et al., 2007), but it is unclear whether refreshing differs from what other WM researchers describe as *retrieval* from LTM or SM. Our preferred account of the strong negative association between recall on WM span tasks and the cognitive load induced by secondary processing tasks is that the extent to which retrieval from SM is involved in WM tasks depends on the amount of time or effort it takes to perform the secondary processing operations.

Levels-of-Processing Effects on WM

One way to assess the extent to which specific WM tasks involve retrieval from LTM is to manipulate a variable known to affect long-term retention. One such variable is LOP; results from the experiments reported by Craik and Tulving (1975) suggested that performance on LTM tasks is highly sensitive to the qualitative level or depth to which memory items are processed when they are initially encoded. It therefore follows that if performance on WM tasks largely reflects retrieval from LTM, manipulation of depth of processing should have a large effect, with deep semantic processing associated with higher levels of WM performance than shallow processing.

However, the results of the few studies that have examined LOP effects on WM tasks are mixed. An early study by Mazuryk and Lockhart (1974) had participants perform an immediate memory task similar to present-day simple and complex WM tasks with conditions that manipulated the depth of processing of the to-be-remembered items. The researchers presented participants with five words for immediate free recall. Participants were instructed to process each word in one of four different ways: either rehearse the word silently, rehearse the word overtly, generate a rhyme (shallow processing), or generate a semantic associate (deep processing). The two rehearsal conditions both produced near-perfect immediate recall, considerably better than performance for the two conditions with a secondary processing demand (rhyme or semantic generation). These latter conditions failed to show an LOP effect in immediate recall, however. After several trials of immediate recall, participants were given a delayed free recall or recognition test on all of the studied words. Semantic processing, despite producing immediate recall performance that was equivalent to phonological processing and worse than either covert and overt rehearsal, resulted in performance superior to all other conditions on both delayed recall and delayed recognition tests. Thus, Mazuryk and Lockhart demonstrated a dissociation between LOP effects on immediate recall in an STM task and delayed recall and recognition.

More recently, the influence of LOP on a novel WM span task—the LOP span task—was reported by Rose et al. (2010). In this task, a target to-be-remembered word was presented (e.g., *bride*, presented in red), with two "processing" words presented immediately afterward (e.g., *dried*, presented in blue, and *groom*, presented in red). The target word could be matched with one of the two processing words based on the color of font (shallow, visual processing), rhyme (intermediate, phonological processing), or meaning (deep, semantic processing). After processing between two and eight of these target matches per list, participants were asked to immediately recall the target words in serial order. Results of three experiments indicated no influence of LOP on immediate recall, but the typical LOP effect was found on a later recognition test. Thus, despite recent theorizing that suggests WM tasks largely involve retrieval from SM, Rose et al. (2010) demonstrated a dissociation between LOP effects on immediate recall in a WM task and a delayed recognition (LTM) test. This pattern was recently replicated in an individual differences study using slightly different procedures (e.g., a final free recall test; Rose, 2011).

Using more traditional WM measures, Loaiza, McCabe, Youngblood, Rose, and Myerson (2011) examined performance on reading and operation span tasks with deep or shallow processing at encoding. For example, for the reading span task, participants judged whether sentences were true or false under two conditions: Half of the sentences required semantic decisions with respect to the to-be-remembered word in capital letters at the end of the sentence (e.g., "The brother of one of your parents is an UNCLE") and the other half of the sentences required a shallow decision (e.g., "A word made up of five letters is UNCLE") Participants were asked to read the entire sentence aloud and respond to the veracity of each statement. After receiving two to five sentences, participants were prompted to recall the final word of each sentence. Then participants performed a distracter task for 2 min, followed by a surprise delayed recall test. In this case, deep processing benefited immediate recall for all list lengths, and Loaiza et al. suggested that, consistent with Unsworth and Engle's (2007) primary–secondary framework, traditional complex span tasks largely measure retrieval from SM.

The purpose of the present experiments was to clarify the circumstances under which LOP effects are and are not found in WM tasks and so shed light on the role of LTM in WM. One procedural difference between the LOP span task and more traditional complex span tasks is that in the latter (e.g., the reading span task), participants are given an orienting question followed by a to-be-remembered word, and the processing decision is made on the to-be-remembered word. In contrast, in the Rose et al. (2010) version of the LOP span task, a to-be-remembered word was presented first, followed by two words that matched the preceding word in color, rhyme, or meaning, and the processing decision was made on the associated words, not the to-be-remembered word. Therefore, it is possible that the lack of LOP effects in Rose et al. (2010) and Rose (2011) was due to the ways in which the procedure diverged from the original LOP procedures used by Craik and Tulving (1975). In the present study, Experiment 1 used the original Craik and Tulving (1975) materials and procedure in an attempt to clarify the conditions under which LOP effects are found in WM tasks. Experiment 2 explored the role of rehearsal and how it might modulate the degree to which LOP affects immediate recall.

Experiment 1

Experiment 1 used the same procedure and stimuli as Craik and Tulving's (1975) Experiment 9. This paradigm is known to produce robust LOP effects on LTM tests. The only difference with the procedure used in the present experiment was that participants performed the visual, phonological, or semantic processing decisions on groups of question–word pairs in the context of a WM test: Immediate recall was required after only a few processing decisions (three or eight), rather than after all of the words were processed. Lists of three or eight items were used to capture the distinction between immediate recall of lists that do (eight) or do not (three) exceed the assumed capacity of the focus of attention (Cowan, 1999). According to Cowan (1999), no more than four chunks of information may be maintained in and reported directly from the focus of attention at any given time; recalling longer lists (sometimes called supraspan lists) will involve retrieval from LTM.

Additionally, the procedure of this novel LOP span task resembles that of the reading span task in that sentences must be read and verified between presentation of each to-be-remembered item. The procedure was as follows: A question was presented (i.e., "Is the following word in UPPERCASE?" for shallow visual processing, "Does the following word RHYME with *X*?" for intermediate phonological processing, or "Is the following word a member of the CATEGORY *X*?" for deep semantic processing, where *X* represents a word that was or was not a rhyme or category member of the following word), and then a to-be-remembered word was presented. Participants were required to answer each question by pressing a key labeled "Yes" or "No" in response to the to-be-remembered word. After each series of either three or eight decisions, participants were asked to recall the to-be-remembered words. All three or eight items of a list were of the same processing type. To compare the LOP effects on WM to those on LTM, following the processing decisions and immediate recall tests, participants solved arithmetic problems for 10 min and then performed a delayed recognition test on all to-be-remembered words from the LOP span task.

Method

Participants and design. Twenty-four Washington University undergraduate students participated in exchange for course credit. All participants were native English speakers. The design was a 3 (LOP: uppercase, rhyme, category) × 2 (list length: 3 or 8 items) × 2 (test: immediate recall, delayed recognition) design. All variables were manipulated within subjects. The main dependent variable was the proportion of words that were correctly recalled on the immediate recall tests and recognized as "old" on the delayed recognition test.

Stimuli. The present study included the same 60 orienting questions and to-be-remembered target words that were used in Experiment 9 of Craik and Tulving (1975), as well as 39 additional questions and target words constructed from the stimuli used in Rose et al. (2010).

Procedure. Participants were tested individually at a desktop computer. The stimuli were presented visually. On each trial, a fixation cross appeared on the monitor where each target word was presented. The participant began each trial by pressing the spacebar when ready, after which an orienting question was displayed for 1,750 ms. After a 250-ms blank screen, a to-be-remembered target word was presented. The participant was instructed to say the word aloud, remember the word for recall at the end of the trial, and press a button labeled "Yes" or "No" in response to the orienting question. The target word remained on the screen until the participant made a response. Prior to testing, the participant was instructed to make each decision as quickly as possible without sacrificing accuracy.

After the processing decision was made, the screen was blank for 750 ms before the next orienting question and target word appeared. At the end of the trial, a green box and a tone cued the participant to recall the target words aloud in the order presented. Participants were told that if they were unable to recall all of the target words, they were to recall as many as possible. Therefore, both serial and free recall scoring procedures could be assessed. Before starting the test trials, participants performed practice trials to familiarize them with the procedure. Recall responses were recorded by electronic voice recorders for later scoring. For the immediate recall test trials, participants performed three trials of three- and eight-item lists for each LOP condition. Trials for the three processing conditions (uppercase, rhyme, category) were mixed in a predetermined random order such that successive trials were not of the same condition. Prior to starting each trial, an instruction screen told the participant the condition on which to base the decision for each word and the questions that preceded each word did as well. After completing all of the immediate recall tests and 10 min of mental arithmetic, participants performed a surprise recognition test.

For the recognition test, the 99 target words that were presented in the LOP span task and 99 new lure words that had never appeared in the experiment were presented individually on the computer monitor. Lures were matched to the target words on the basis of length and word frequency. For each word, participants were instructed to indicate whether that word was "old," meaning it was to be read aloud and remembered on one of the immediate recall tests, or "new," meaning the word was never presented in the experiment.

Results

It was first verified that participants performed the processing operations of the LOP span task. The proportion of correct processing decisions was high in all conditions: visual = .99 (SD = 0.01), phonological = .92 (SD = 0.06), and semantic = .97 (SD = 0.05). The main effect was significant, $F(2, 46) = 20.75$, $p < .001$, because the proportion of correct processing decisions was higher in the shallowest (visual) condition than in both the phonological and semantic conditions, $t(23)\text{s} > 2.73$, $p\text{s} < .05$. Reaction times were also faster for the visual condition (M = 1,184 ms, SD = 378) than for both the phonological (M = 1,459 ms, SD = 418) and semantic (M = 1,420 ms, SD = 446) conditions, $t(23)\text{s} > 7.62$, $p\text{s} < .001$.

The proportion of words recalled irrespective of serial position on the immediate recall tests of the LOP span task are presented in the upper half of Table 1 (the pattern of results for serial recall scoring was the same). As can be seen, deep (semantic) LOP did

Table 1
Mean (SEM) Proportion of Items Correctly Recalled on the LOP Span Task and Correctly Recognized as Old on the Delayed Recognition Test for Items Initially From Three- or Eight-Item Lists

	Level of processing		
Task	Visual	Phonological	Semantic
Immediate Recall			
3 items	.99 (.01)	.92 (.02)	.98 (.01)
8 items	.56 (.02)	.44 (.03)	.51 (.02)
Delayed Recognition			
3 items	.61 (.05)	.66 (.04)	.73 (.04)
8 items	.69 (.04)	.69 (.03)	.82 (.03)

Note. The false alarm rate was .19 (.02).

not benefit immediate recall relative to the shallowest (visual) LOP, even for the longer eight-item lists. In contrast, delayed recognition of words that were initially processed in the LOP span task did demonstrate a benefit of deeper LOP. The mean proportions of words correctly recognized as old are presented in the bottom half of Table 1. The immediate recall and delayed recognition data were submitted to separate 3 (LOP) × 2 (list length) repeated-measures analyses of variance (ANOVAs). For immediate recall, the effect of LOP was significant, $F(2, 46) = 20.9$, $p < .001$, but it was not as predicted by the LOP framework: The shallowest (uppercase) processing condition ($M = 0.77$) was significantly better than both the rhyme ($M = 0.68$), $F(1, 23) = 33.7$, $p < .001$, and the category ($M = 0.75$) processing conditions, $F(1, 23) = 4.2$, $p = .05$. As expected, there was a main effect of list length such that a greater proportion of words were recalled from three-item lists than eight-item lists, $F(1, 23) = 753.9$, $p < .001$. However, list length did not interact with LOP, $F(2, 46) = 2.2$, $p = .12$.[1]

For delayed recognition, the effect of LOP was significant, $F(2, 46) = 11.8$, $p < .001$, because semantically processed words were recognized better than phonologically or visually processed words. There was also a main effect of list length, $F(1, 23) = 10.3$, $p < .01$, as words initially presented in eight-item lists were better recognized than words presented in three-item lists. LOP and list length did not interact, $F(2, 46) = 0.8$, $p = .47$.

The data depicted in Figure 1 illustrate the dissociation between LOP effects on the immediate and delayed memory tests. The immediate recall data on the left are the average proportions of words recalled in the case, rhyme, and category LOP conditions, collapsed across list length. The delayed recognition data on the right side of the figure are the average proportions of words recognized as old for the case, rhyme, and category LOP conditions, collapsed across list length. The comparison between the shallowest (visual) LOP and deepest (semantic) LOP conditions is of particular interest. Immediate recall was significantly better for the visual than the semantic condition, but delayed recognition showed a 13% advantage of semantic over visual processing.

Discussion

The results of Experiment 1 demonstrate that, despite the use of the same LOP paradigm used by Craik and Tulving (1975), the effect was eliminated by testing memory after only a few decisions, as opposed to after all of the decisions on a delayed test. These findings are consistent with those of Rose et al. (2010) and Rose (2011) and point to a striking dissociation between WM and LTM. They are, however, inconsistent with the findings of Loaiza et al. (2011) and may appear contrary to recent theorizing about WM, which suggests that performance on WM tests involves retrieving items from LTM. Thus, an important theoretical question arises: Why, if retrieval from LTM was involved, did deeper LOP fail to affect WM performance?

Because the WM tests involved maintenance of a relatively small set of information over short retention intervals, whereas the LTM test involved retention of a much larger set of material over a long interval, different processes were likely involved in performance on the two types of tests. One difference between the WM and LTM tests concerns intentional encoding and active maintenance of the to-be-remembered words. The WM tests required that the words be encoded and maintained for immediate recall. Participants likely tried to maintain the words by covertly retrieving them between performance of the LOP decisions and presentation of subsequent stimuli (McCabe, 2008). Such covert retrievals may have served to return the to-be-remembered words to the focus-of-attention so that, at the time of retrieval, the words were accessible. Thus, the WM tests likely involved either reporting items directly from the focus-of-attention or retrieving items from SM that were highly activated because they were recently refreshed in the focus-of-attention when the participant covertly retrieved them. Whereas maintenance of words for immediate recall probably involved the use of an articulatory code, retrieval from LTM after some minutes likely relied on semantic codes. The results may have differed from those of Loaiza et al. (2011) because the secondary tasks in more traditional span tasks take longer to perform and are more difficult than the secondary tasks in the LOP span task. In sum, the pattern of LOP effects on WM and LTM likely differed because the nature of WM retrieval in Experiment 1 was different from the type of retrieval involved in the LTM test, at least for those particular task conditions.

If, however, an immediate test was unexpected on the LOP span task, there would be no need for participants to maintain the words actively in preparation for recall. Rather, participants would incidentally encode the words according to the LOP condition and, as participants process more and more information, previous items would no longer be rehearsed and so they would be displaced from the focus of attention. It follows that if an immediate recall test was administered after the items had already been processed, recalling the items on a surprise test would require retrieving them from LTM, since they were no longer maintained in PM. The type of retrieval would therefore be similar to that of a surprise LTM test administered after a filled retention interval, even though recall is

---

[1] LOP did interact with serial position, but this was only true for eight-item lists, $F(14, 322) = 2.5$, $p < .05$; three-item lists: $F(4, 93) = 1.8$, $p = .13$. Pairwise comparisons revealed that this was due to recall being higher for the visual condition than both the phonological and semantic conditions for serial Positions 1 and 2 ($p < .05$), lower for the phonological condition than the semantic condition for Position 4 ($p < .05$), and lower for the phonological condition than both the semantic and visual conditions for Position 5 ($p < .05$). Differences between conditions were not observed for any other serial position.

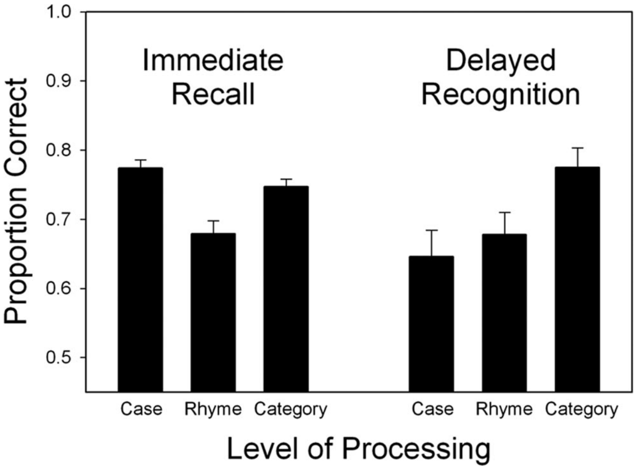

Figure 1. Mean proportion of words recalled on the immediate tests of the levels-of-processing span task (collapsed across list length) and subsequently recognized as old target words on the delayed recognition test as a function of levels-of-processing. Case = shallow, visual processing; rhyme = phonological processing; category = deep, semantic processing. Error bars represent the standard error of the mean.

immediate. As a result, deeper LOP should benefit retrieval on a surprise immediate recall test. Experiment 2 was designed to test this hypothesis.

Experiment 2

The same general procedure used in Experiment 1 was used in Experiment 2, except that participants were not expecting the immediate recall test. The participants were told that we were interested in how fast uppercase, rhyme, and category decisions could be made, and so they were to make each processing decision as quickly and accurately as possible. However, just one surprise immediate recall test was administered to each participant. After processing the final series of words, each participant was asked to recall as many words as possible from that series. Informal post-experimental questioning suggested that participants truly were not expecting this test. LOP and list length were manipulated between subjects for the surprise immediate recall test. Participants were randomly assigned to one of three processing conditions: One third of the participants were required to recall visually processed items (from a series that was three items long for half of these participants and eight items long for the other half); another third recalled phonologically processed items (either three or eight) and the final third recalled semantically processed items (either three or eight).

Because Experiment 2 assessed immediate recall when testing was not expected, it seemed unlikely that participants would actively maintain the words. Accordingly, it was predicted that immediate recall following incidental encoding would demonstrate a benefit of deep processing. However, because the extent to which retrieval from LTM is involved in immediate recall depends on list length (Unsworth & Engle, 2006), it was also predicted that there would be a tradeoff between deep and shallow processing as a function of list length. In the shallow processing condition, attending to the case of a word's font is unlikely to produce much interference with representing to-be-remembered words. According to Nairne (1990), quickly attending to the case of a word should not cause as much feature overwriting as attending to the word's phonological or semantic characteristics, and so recalling three words after case decisions should be easier than after rhyme or semantic decisions. Moreover, the visual processing condition involved processing the same orienting question each time ("Is the following word in uppercase?"), whereas the phonological and semantic conditions involved reading a novel sentence for each item and a word that was phonologically or semantically associated to a word to be recalled on the surprise test. Because the rhyme and semantic associates share many features with the to-be-recalled target words, they may interfere with the retrieval process (Nairne, 1990). Thus, shallow processing should be best for recovery of three-item lists because there is minimal interference from the secondary processing operation. In contrast, deep processing should be best for recovery of the eight-item lists because recall would principally rely on retrieval from LTM.

Method

Participants and design. Forty-eight undergraduate students participated in exchange for course credit. The design was a 3 (LOP: uppercase, rhyme, category) × 2 (list length: three items, eight items) design. LOP and list-length variables were between-subjects factors. The dependent variable was the proportion of words recalled on the surprise immediate recall test.

Procedure. Participants were instructed to make each processing decision as rapidly and accurately as possible. Following each set of three or eight decisions, a green box appeared. Participants were instructed to pause until the next trial began. The duration of the pause was set to the mean duration that participants took to recall three- or eight-item lists for the LOP span task in Experiment 1 (approximately 3.5 s and 10.5 s for three- and eight-item lists, respectively). On the last trial, when the green box was displayed, an additional set of instructions appeared on the screen, which read, "Please repeat the words you said aloud on this trial. Try to remember as many as you can."

Results

We first verified that participants performed the processing operations of the LOP span task. The proportion of correct processing decisions was high in all conditions: visual = .99 (SD = 0.02), phonological = .96 (SD = 0.04), and semantic = .95 (SD = 0.07). The main effect was significant, $F(2, 94) = 13.69$, $p < .001$, because the proportion of correct processing decisions was higher in the shallowest (visual) condition than in both the phonological and semantic conditions, $t(47)$s > 4.6, ps $< .001$. Reaction times were also faster for the visual condition (M = 688 ms, SD = 226) than for both the phonological (M = 817 ms, SD = 218) and semantic (M = 850 ms, SD = 229) conditions, $t(47)$s > 9.49, ps $<$ .001, and the phonological condition was faster than the semantic condition, $t(47) = 2.78$, $p < .01$.

The mean proportions of words recalled on the surprise recall test are presented in Figure 2, alongside the immediate recall results of Experiment 1. As predicted, an LOP effect was obtained on the immediate recall tests but only for the eight-item lists. That is, the deepest LOP was best for the supraspan list. In contrast, the shallowest LOP was best for the three-item list. These observations

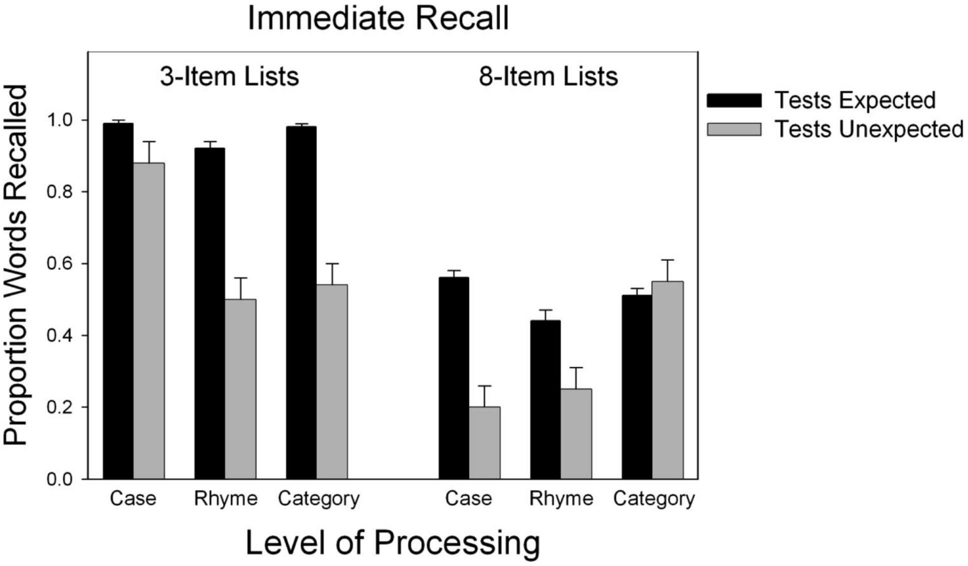

Figure 2. Mean proportion of words recalled on the immediate tests for Experiment 1 (when immediate tests were expected) and Experiment 2 (when immediate tests were unexpected) as a function of level of processing. Case = shallow, visual processing; rhyme = phonological processing; category = deep, semantic processing. Error bars represent the standard error of the mean.

were statistically confirmed by an ANOVA with LOP and list length as between-subjects factors. The effect of LOP was significant, $F(2, 42) = 4.7$, $p < .05$. As expected, there was also a main effect of list length such that a greater proportion of words were recalled from three-item lists than eight-item lists, $F(1, 42) = 35.5$, $p < .001$. In addition, list length interacted with LOP, $F(2, 42) = 14.8$, $p < .001$, because the deepest LOP benefited recall of items from the eight-item lists, $F(2, 42) = 10.7$, $p < .01$, whereas the shallowest LOP benefited recall of items from the three-item lists, $F(2, 42) = 8.8$, $p < .001$.

Discussion

As predicted, a benefit of deeper LOP was obtained on immediate recall, but only for the eight-item lists. This finding is particularly interesting when considered alongside the results of Experiment 1. Consider, for example, the difference in LOP effects on immediate recall between Experiments 1 and 2 (see Figure 2). Even though participants made the same processing decisions on the same words in both experiments, immediate recall did not show a benefit of deeper LOP in Experiment 1, but it did in Experiment 2, specifically for the supraspan (eight-item) lists.

Direct comparisons should be treated with caution due to the methodological differences between Experiments 1 and 2. However, the point is that in Experiment 1 participants knew of the upcoming immediate recall test on each trial, so they were likely trying to maintain the target words; whereas, in Experiment 2, participants were not expecting a recall test, so they would not have maintained the words. As a result, LOP did not affect immediate recall in Experiment 1 but did affect immediate recall in Experiment 2. This finding is consistent with the hypothesis that when active maintenance processes are eliminated, immediate recall on a WM task demonstrates an LOP effect similar to LTM tasks, for supraspan lists at least.

General Discussion

In the present study, Craik and Tulving's (1975) original materials and procedure were used in a complex WM span task to assess the effect of LOP on WM. Although this procedure is known to have robust effects on LTM, deep LOP at encoding did not benefit WM in Experiment 1 but did benefit immediate recall of eight-item lists on a surprise test in Experiment 2. We argue that the reason for this pattern of findings is that in Experiment 1 participants encoded to-be-remembered items intentionally, and presumably they attempted to actively maintain those items in preparation for an immediate memory test. For the surprise immediate recall test in Experiment 2, however, participants saw no need to actively maintain the words. For recall of three-item lists in the second experiment, shallow processing resulted in the best level of recall, most likely because this condition produced the least interference with the target items (case decisions were significantly faster and more accurate than rhyme or category judgments, which suggests shallow processing was easier and less distracting than deeper processing). For recall of eight-item lists, it seems probable that deeper processing was best because recall depended largely on cue-driven retrieval from long-term (secondary) memory (Craik & Lockhart, 1972; Rose et al., 2010; Unsworth & Engle, 2007).

Taken together, these results suggest that the reason deep LOP did not benefit immediate recall relative to shallow LOP in Experiment 1 was because participants were actively retrieving (or refreshing) the to-be-remembered items between the processing phases of the task to maintain them in PM, consistent with the time-based resource sharing model (Barrouillet et al., 2004) and McCabe's (2008) covert retrieval model of performance on complex span tasks. When active maintenance of the to-be-remembered items was eliminated, because immediate recall tests were unexpected, recall of eight-item lists would have depended on retrieving the items from SM (Unsworth & Engle, 2007). Thus the initial encoding of deep semantic cues facilitated later recall (Craik & Lockhart, 1972).[2]

Levels-of-Processing Effects on WM

The results of Experiment 1 conceptually replicate the pattern of results from Rose et al. (2010) and Rose (2011) and extend them using different procedures (e.g., with sentence stems similar to the reading span task). The results of Experiment 1 differ from those of Loaiza et al. (2011), which showed that deeper LOP did benefit immediate recall on more traditional complex span tasks (e.g., the reading span task). What might account for this difference?

As discussed in the introduction, Unsworth and Engle's (2007) PM–SM framework assumes that during the performance of complex span tasks, when a participant switches attention away from encoding to-be-remembered items to perform the processing task (e.g., reading sentences and judging their veracity), the to-be-remembered items are displaced from PM, and so they must be retrieved from SM. As suggested by Unsworth and Engle's (2007) PM–SM framework, McCabe's (2008) covert retrieval model of performance on complex span tasks posits that after to-be-remembered items have been displaced from PM, participants try to retrieve or refresh the items in between the processing phases of the task. Findings that show LTM is better for items initially retrieved from complex versus simple span tasks (McCabe, 2008; Rose, 2011) or from longer versus shorter lists (Rose et al., 2010) suggest that these covert retrievals provide distributed practice at retrieving the items from SM, which involves deeper and more elaborate retrieval operations than does reporting items from PM (Craik, 1970).

However, the PM–SM framework does not distinguish between types of secondary task that are more or less likely to displace items from the focus of attention. According to Unsworth and Engle (2007), "if attention is removed, because new information is intentionally being processed or because attention has been captured by environmental stimuli (e.g., a flashing light), representations are displaced from PM" (p. 107). A strict prediction derived from this hypothesis is that as soon as the focus of attention shifts to performing other cognitive operations, returning the to-be-remembered items to PM requires retrieval from SM.

Other researchers have suggested that when attention shifts from maintenance of memory items to secondary processing operations in span tasks, there is a more subtle shift in the nature of processes involved. For example, the time-based resource-sharing model of Barrouillet and colleagues suggested that span tasks vary in the amount of distraction produced by the secondary task (Barrouillet et al., 2004; Barrouillet et al., 2011). Like McCabe (2008), Barrouillet et al. (2011) assumed that participants attempt to refresh to-be-remembered items between the processing and retrieval phases of complex span tasks, and that this "attentional refreshing" can be carried out along with easy secondary tasks (such as repeatedly articulating the same word, i.e., articulatory suppression) that are minimally demanding of attention. In contrast, more difficult secondary tasks prevent one from attentionally refreshing items and so performance on WM span tasks declines. Barrouillet et al. (2011) have shown that there is a very systematic decline in WM span as a function of the amount of time that attention is drawn away from maintaining to-be-remembered items. One way to interpret this pattern, which Barrouillet et al. (2011) recently suggested is a law, is that the extent to which items must be retrieved from SM at the time of recall depends upon the amount of time that the focus of attention has been drawn away from maintaining to-be-remembered items. It is not that to-be-remembered items are totally displaced from the focus of attention as soon as it shifts to the secondary task. Rather, complex span tasks that involve lengthy, difficult processing phases should involve retrieval from SM to a greater extent than tasks with shorter or easier processing tasks.

Thus, according to this hypothesis, the extent to which LOP effects should appear on WM tasks depends on the amount of time and difficulty of the secondary processing tasks. Specifically, LOP effects should appear in WM tasks when the processing tasks are more difficult and time-consuming (as was the case in Loaiza et al., 2011, in which reading and verifying sentences or math operations took approximately 5,000 ms per item) than when the processing tasks are easier and shorter (as was the case in Rose et al., 2010; Rose, 2011; and Experiment 1 of the current study, in which the LOP decisions took closer to 1,000 ms per item), particularly if the shallow processing task is more time-consuming than the deep processing task.

To test this idea, we examined the relation between LOP effects on WM and amount of processing time by comparing deep and shallow processing conditions from all of the experiments that have manipulated LOP in a complex WM span task in terms of immediate recall performance and the duration of the processing phase. These data, aggregated from a total of eight independent experiments on a total of 322 participants, are presented in Table 2. The size of the LOP effect (the difference in recall accuracy between deep and shallow conditions) on WM is plotted as a function of the difference in the average processing decision time between the deep and shallow conditions in Figure 3. As may be seen, there is a strong relation between the size of the LOP effect and the difference in processing times between deep and shallow conditions ($r = -.63$, $p < .05$). This suggests that LOP effects on WM depend in part on the amount of time required to perform the secondary processing operations, a pattern that is consistent with the notion that the extent to which LTM is involved in WM tasks depends on the amount of time that attention is focused on the processing phase of the task.

Put another way, if secondary processing times are short, items can be maintained in PM and the LOP manipulation will have little effect. As secondary processing tasks become more difficult, processing times lengthen, the necessity to retrieve from SM increases, and the size of the LOP effect will therefore also increase. Performance will be particularly poor for the combination of long processing times and shallow processing, and in such cases the

---

[2] Similarly, Speer, Jacoby, and Braver (2003) showed that biasing the use of maintenance-focused versus retrieval-focused processes by manipulating participants' test expectations (immediate recognition of four vs. eight words) led to unique brain areas being activated (in addition to several overlapping areas), even though the overall memory task was identical (immediate recognition of six words). That test expectations can change how a task is performed is consistent with the processing approach to the WM–LTM distinction, as well as the dual-component model of WM (Unsworth & Engle, 2007). We thank an anonymous reviewer for suggesting this point.

Table 2
Average Processing Decision Times for Deep and Shallow Processing Conditions From Studies on LOP Effects on Working Memory Span Tasks

	Response time (ms)			Immediate recall (%)		
Experiment	Deep	Shallow	Difference	Deep	Shallow	Difference
Rose et al. (2010), Exp. 1a	1,243	630	613	83	83	00
Rose et al. (2010), Exp. 1b	1,143	596	547	81	79	02
Rose et al. (2010), Exp. 2	1,140	1,852	−712	70	66	04
Loaiza et al. (2011), Exp. 1a	5,190	5,570	−380	62	58	04
Loaiza et al. (2011), Exp. 1b	5,173	5,349	−176	69	63	06
Loaiza et al. (2011), Exp. 2	3,518	3,530	−12	62	57	05
Rose (2011)	1,499	1,483	16	72	72	00
Rose & Craik (current study), Exp. 1	1,420	1,184	236	75	77	−03

LOP effect will be greatest; this is the situation shown in the upper left-hand quadrant of Figure 3.

On the Distinction Between WM and LTM: A Processing Approach

Taken together, the present findings have implications for how memory theories should conceptualize the distinction between WM and LTM. In particular, they resolve previous findings (i.e., Experiment 1; Rose et al., 2010; Rose, 2011), which showed that immediate recall on a WM test did not show an LOP effect—a result that appeared to be inconsistent with the hypothesis that WM tests involve retrieval from LTM (e.g., Unsworth & Engle, 2007). However, the present findings show that performance on WM and LTM tests can demonstrate both similarities and differences depending on various features of the tasks (e.g., involvement of active maintenance processes, test expectations, list length, etc.). Thus, although it is unlikely that retrieval over the short- and long-term depend on entirely different memory systems, it is also unlikely that WM and LTM represent identical constructs. Rather, WM and LTM tests likely have both shared and unique processes, and the extent to which WM and LTM tests demonstrate similar principles depends on the extent to which the two types of tests invoke similar encoding, retention and retrieval processes (Rose et al., 2010; Unsworth, 2010; see also, Craik & Lockhart, 1972, for a similar discussion on the STM–LTM distinction).

With regard to other researchers' conceptualization of the WM–LTM distinction, the critical difference between Baddeley's (2003) model and embedded process models (e.g., Cowan, 1999, 2005; Oberauer, 2010) appears to be a preference for structural as opposed to processing concepts. Baddeley (2003) argued that the relation between WM and LTM necessitates a multimodal "episodic buffer" to which information is transferred (see p. 836). To us, however, the idea of a separate structure that items are "downloaded into" from LTM seems unnecessary psychologically and somewhat implausible neurologically. In contrast, embedded process models agree that when items have been displaced from PM

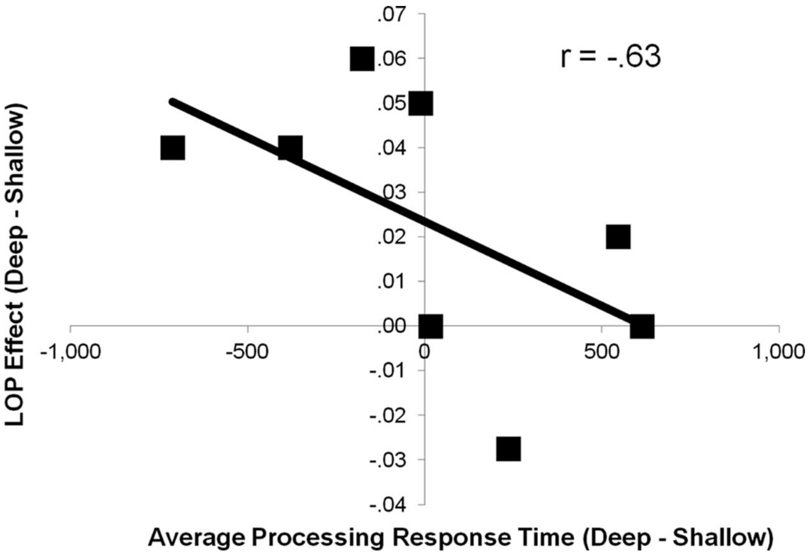

Figure 3. Levels-of-processing (LOP) effects on immediate recall (deep – shallow) as a function of the difference in processing decision times between the deep and shallow processing conditions from eight experiments that manipulated levels of processing on working memory span tasks ($N = 322$).

(or the focus of attention), returning them back to PM requires retrieving their representations from activated long-term or secondary memory. Some prefer to use the term retrieval to describe this process (Cowan, 2005; McCabe, 2008; Oberauer, 2010; Rose et al., 2010; Unsworth & Engle, 2007), but the terms attentional refreshing (Barrouillet et al., 2004; Johnson, 1992), attentional focusing (McElree, 2006), and focus switching (Zhang & Verhaeghen, 2009) have all been used as well. It is unclear whether these differences are simply a matter of terminology or whether the different terms reflect fundamental differences in the nature of the process. Clarification of this point is an important goal for future work. A more specific question is whether attentionally refreshing information in WM is any different from retrieval of information from the activated portion of LTM.

In our opinion, it is not strictly correct to state that information "in WM" represents information in a separate temporary store (e.g., Baddeley, 2003) or perhaps even in the activated subset of LTM (e.g., Cowan, 2005; Oberauer, 2010). Rather, WM performance involves a mixture of both PM and SM representational codes, depending on whether it is possible to continuously attend to PM codes, which in turn largely depends on the presence and difficulty of secondary task operations. For verbal materials, these PM and SM codes are likely to be phonological/articulatory and semantic/conceptual, respectively, in which case conscious maintenance of target items in WM does not represent activation of LTM but rather reflects attending to different features of representations. Thus, it may be preferable to talk about information "maintained in PM" or "retrieved from SM" as attention paid to the particular representational codes involved in the task at hand—for example, articulatory operations in a simple span task with few verbal items, a mixture of articulatory and semantic codes in complex span tasks in which some items must be retrieved from SM, and mostly semantic codes in complex WM tasks in which difficult secondary task operations necessitate SM retrieval for most items (see also Craik & Levy, 1970).

In conclusion, the current study considered the distinction between short- and long-term retention by examining LOP effects on WM and LTM tasks. A review of the literature, the results of two experiments, and a joint analysis of studies examining LOP effects on WM suggest that there are both similarities and differences between LOP effects on WM and LTM. To account for this pattern, a processing approach to the WM–LTM distinction was advanced. Rather than viewing WM and LTM as representing distinct systems, this account suggests that similarities and differences between WM and LTM depend on the extent to which there is a match or mismatch between the encoding, maintenance, and/or retrieval processes involved in performance on the tasks used to assess WM and LTM.

References

Baddeley, A., & Warrington, E. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 176–189. doi:10.1016/S0022-5371(70)80048-2

Baddeley, A. D. (1966a). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology, 18,* 362–365. doi:10.1080/14640746608400055

Baddeley, A. D. (1966b). The influence of acoustic and semantic similarity on long-term memory for word sequences. *The Quarterly Journal of Experimental Psychology, 18,* 302–309. doi:10.1080/14640746608400047

Baddeley, A. D. (1986). *Working memory.* New York, NY: Clarendon Press/Oxford University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829–839. doi:10.1038/nrn1201

Baddeley, A. D. (2007). *Working memory, thought and action.* Oxford, England: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York, NY: Academic Press.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133,* 83–100. doi:10.1037/0096-3445.133.1.83

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review, 118,* 175–192. doi:10.1037/a0022324

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York, NY: Cambridge University Press.

Cowan, N. (2005). *Working memory capacity.* New York, NY: Psychology Press. doi:10.4324/9780203342398

Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning & Verbal Behavior, 9,* 143–148. doi:10.1016/S0022-5371(70)80042-1

Craik, F. I. M., & Levy, B. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology, 86,* 77–82. doi:10.1037/h0029799

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11,* 671–684. doi:10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294. doi:10.1037/0096-3445.104.3.268

Johnson, M. K. (1992). MEM: Mechanisms of recollection. *Journal of Cognitive Neuroscience, 4,* 268–280. doi:10.1162/jocn.1992.4.3.268

Jonides, J., Lewis, R., Nee, D., Lustig, C., Berman, M., & Moore, K. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59,* 193–224. doi:10.1146/annurev.psych.59.103006.093615

Loaiza, V., McCabe, D., Youngblood, J., Rose, N. S., & Myerson, J. (2011). The influence of levels of processing on recall from working memory and delayed recall tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1258–1263. doi:10.1037/a0023923

Mazuryk, G. F., & Lockhart, R. S. (1974). Negative recency and levels of processing in free recall. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 28,* 114–123. doi:10.1037/h0081971

McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests. *Journal of Memory and Language, 58,* 480–494. doi:10.1016/j.jml.2007.04.004

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 155–200). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(06)46005-9

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18, 3,* 251–269. doi:10.3758/BF03213879

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421. doi:10.1037/0278-7393.28.3.411

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation, 51,* 45–100. doi:10.1016/S0079-7421(09)51002-X

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences, 9,* 374–380. doi:10.1016/j.tics.2005.06.009

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex, 43,* 135–145. doi:10.1016/S0010-9452(08)70451-9

Rose, N. S. (2011). *Individual differences in working memory, secondary memory, and fluid intelligence: Evidence from the levels-of-processing span task.* Manuscript submitted for publication.

Rose, N. S., Myerson, J., Roediger, H. L., & Hale, S. (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 471–483. doi:10.1037/a0018405

Rose, N. S., Olsen, R. K., Craik, F. I. M., & Rosenbaum, R. S. (2011). Working memory and amnesia: The role of stimulus novelty. *Neuropsychologia.* Advance online publication. doi:10.1016/j.neuropsychologia.2011.10.016

Shallice, T., & Warrington, E. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly Journal of Experimental Psychology, 22,* 261–273. doi:10.1080/00335557043000203

Speer, N. K., Jacoby, L. L., & Braver, T. S. (2003). Strategy-dependent changes in memory: Effects on behavior and brain activity. *Cognitive, Affective, & Behavioral Neuroscience, 3*(3), 155–167.

Thorn, A., & Page, M. (Eds.). (2008). *Interactions between short-term and long-term memory in the verbal domain.* Hove, United Kingdom: Psychology Press.

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica, 134,* 16–28. doi:10.1016/j.actpsy.2009.11.010

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54,* 68–80. doi:10.1016/j.jml.2005.06.003

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114, 1,* 104–132. doi:10.1037/0033-295X.114.1.104

Waugh, N. C., & Norman, D. (1965). Primary memory. *Psychological Review, 72,* 89–104. doi:10.1037/h0021797

Zhang, Y., & Verhaeghen, P. (2009). Glimpses of a one-speed mind: Focus-switching and search for verbal and visual, and easy and difficult items in working memory. *Acta Psychologica, 131,* 235–244. doi:10.1016/j.actpsy.2009.05.009

Received June 27, 2011
Revision received December 12, 2011
Accepted December 13, 2011 ■